Meaningless Dictionaries

Maarten Janssen

ILTEC

The creation of word meaning is one of the most time consuming parts of creating a dictionary. Although it is commonly thought that providing definitions for words is the primary function of dictionaries, it is not the most frequent one. Most dictionaries are used for looking up much more basic information, such as to see whether a word exists or to see whether it is spelled correctly. Dictionaries are relatively good at providing complete definitions for individual words but are not necessarily well equipped for more basic tasks. For many of these smaller tasks, users would be better off using smaller databases-or dictionaries-that focus only on the information the user is looking for rather than searching in a general language dictionary. A dictionary that leaves out most of the details traditionally included in the lexical entry not only makes it easier for the user to find the information he is looking for but also allows the lexicographer to put more focus on the relevant data. It does this by focussing on a single type of information; it becomes more feasible to treat it completely, consistently and coherently for the entire lexicon. The Open Source Lexical Information Network-henceforth OSLIN-is an attempt to create such singe-task lexical resources. This paper explains both the advantages and problems of such an approach.

1. Introduction

Defining word meanings is probably the most problematic part of making dictionaries. Not only do semantic definitions make up the bulk of the text in a dictionary, and is the process of writing them hence necessarily time-consuming, it is also a task pestered by a wide array of practical and fundamental problems. To name but some of the issues: (1) it is difficult to establish the correct set of word-senses for a word, and different dictionaries of the same language differ significantly in the set of senses they give for any given word; (2) dictionary definitions do not really provide necessary and sufficient conditions for the concepts they attempt to define (although they are still more successful than any alternative attempts at word definitions); (3) it is virtually impossible to keep the structure and style of the definitions fully consistent throughout the entire dictionary and it is hard not to use words that are not themselves in the dictionary. For these and other reasons, several authors have recently argued that word senses do not exists in the first place (Kilgarriff 1997, Hanks 2000).

Since dictionaries are reference works that provide meanings for words, the creation of word-sense definitions is an inevitable obligation. However, as is well known, looking up word meanings might be the most characteristic purpose of dictionaries, but it is not the most frequent one. Most dictionary uses are for looking up much more basic information, such as whether a word exists, and whether or not it is spelled correctly. And while dictionaries are (relatively) good at providing complete definition for individual words, they are not necessarily well equipped for such more basic tasks.

For many of these smaller tasks, users would be better server by smaller lexical resources (or dictionaries) focusing only on the information the user is looking for. A dictionary that leaves out most of the details that are traditionally included in the lexical entry not only makes it easier for the user to find the information (s)he is looking for, but also allows the lexicographer to put more focus on the specific type of data the dictionary is intended for: by concentrating on a single type of information, it becomes more feasible to treat that information completely, consistently and coherently for the entire lexicon. The information that should hence be most prominently excluded from such resources is the semantic definition of the entries.

The *Open Source Lexical Information Network* (Janssen 2005a; henceforth OSLIN) is an attempt to create such singe-task lexical resources. The set-up of OSLIN allows these lexical resources to be built and maintained separately, yet despite their relative independence, all separate resources are organized together in a network of lexical information. The set up of OSLIN is in principle language-independent, but currently only Portuguese data are under development. For the Portuguese OSLIN, the team at ILTEC managed to create a number of high-quality dedicated dictionaries, within a relatively small span of time and with limited means. The resulting databases can be accessed freely via a web site called *Portal da Língua Portuguesa*.

Despite the obvious advantages for both the user and the lexicographer of creating lexical resources without semantic definition, there are also at least two serious drawback in doing so. Firstly, many of the seemingly meaning independent phenomena are still partly word sense dependent, leading to problems in several of the dedicated dictionaries. And secondly, the users of the site are often confused by the absence of word meaning. In this paper, we try to demonstrate the advantages of dedicated mini-dictionaries, as well as the problems the approach generates, including the solutions chosen for these problems in the set-up of OSLIN. Some of the dictionaries that have been built for the *Portal* thus far will be used to illustrate both the advantages and the problems. Before turning to the discussion of the meaningless dictionaries, the next section will give a brief introduction to the *Portal* web site and the design of OSLIN.

2. *Portal da Língua Portuguesa*

The *Portal da Língua Portuguesa* is a freely accessible web site, oriented towards the general public, providing information concerning the Portuguese language. Its primary function is to provide formal characteristics of the words of the lexicon of the language, such as spelling, pronunciation, inflection, derivations etc. The web site in principle concerns all national variants of Portuguese, although it has an emphasis on the European Portuguese variety.

The lexical data of the *Portal* are extracted directly from the OSLIN databases for Portuguese, with the MorDebe lexicon as its core. The Portuguese MorDebe is a lexicographically controlled lexicon, with currently around 130,000 lexical entries and over 1,3 million word forms. New lexical entries are constantly being added. It is the largest open source lexicon for Portuguese, and is more and more commonly cited as one of the reference resources for Portuguese.

Apart from the MorDebe lexicon, the database contains a number of single-task dictionaries, all of which are linked to the main lexicon. At this moment, the Portal provides a dictionary of deverbal nouns, a dictionary of gentiles and toponyms, and a dictionary of loanwords. Apart from these dictionaries that are currently available, a dictionary of syllabic division, a dictionary of verbal and nominal inflection, and a dictionary of pronunciation are almost ready to be launched, and several others are still under development. All databases, as well as the web site itself, were developed by the ILTEC institute in Lisbon.

There are two modes of browsing lexical information on the *Portal*. The first way is to look at the complete record for each individual word. The complete record will not only show the word itself, but also collect all the information concerning the word from all the separate databases of OSLIN such as it grammatical class, its inflectional paradigm, and its derivational relations. The second way of browsing information is via the dedicated dictionaries themselves, which only display the information contained within that dictionary (although each record of the dictionary will point back to the complete record for the individual word). For instance, the loanword dictionary contains the information that *meeting* is a loanword from English, and what its equivalent in Portuguese is. But the information about its status of being a loanword is displayed in the main MorDebe page for that word as well, as can be seen in figure 1.

Figure 1. The dual access mechanism of OSLIN

The *Portal* explicitly invites users to provide feedback if they encounter any information in the database they think might be incorrect. Large-scale databases inevitably contain errors, and finding those is not a trivial task. Therefore, by asking for the active participation of the user base, it becomes possible to improve OSLIN by carefully verifying any potential errors pointed out by the users of the *Portal*.

The Portal does not only feature the OSLIN lexical data, but also has some other resources that are relevant for the common speaker of Portuguese. The most prominent ones at the web site at this moment are the *acordos ortográficos* (official spelling) and a dictionary of linguistic terms. Several other resources that might be useful for the language user are currently under development – but progress is slow due to the small budget of the project.

The Portal is a relatively popular web site, with around 2,500 visitors each day, in spite of the little effort on marketing so far. The website, and the underlying databases, are also used for several academic courses, as a reference source by the *Ciberdúvidas* language consultancy, and by various teachers for the preparation of their classes. The hope is that the number of visitors will significantly grow after the marketing campaign we are currently undertaking.

3. Dedicated meaningless dictionaries

This chapter tries to illustrate the advantages of dedicated dictionary without full meaning descriptions, using three dictionaries from OSLIN as an example: firstly, the MorDebe database is the basic lexicon of OSLIN, which provides information regarding orthography and inflection. Secondly, the dictionary of deverbal nouns, which provides the nominal form of each verb in MorDebe. And thirdly, the dictionary of gentiles, which provides the adjectival form related to a large number of toponyms.

The dictionaries discussed here are not the only ones based on the idea of meaningless dictionaries. Firstly, there are other meaningless dictionaries in OSLIN, either already finished or currently under development, such as a database of syllabic division, a dictionary of deadjectival nouns, a pronunciation dictionary, and a dictionary of loanwords. Secondly, many traditional dictionaries provide small lists of specific words without meaning definitions at the beginning or the end of the book—the Houaiss dictionary contains a list of all sounds of animals, such as

barking for *dog*—; the LDOCE dictionary provides a list of gentiles for the countries around the world; the DLPC also includes the names of all currencies in that list; and of course all the official vocabularies, such as the Dutch *Woordenlijst der Nederlandse Taal*, and in a sense even the thesauri and reverse dictionaries are lexica without meaning definitions as well. The fact that there are so many word-lists of this type available is in and by itself an indication that they serve a purpose. In this chapter we attempt to look at the advantages and disadvantages of meaningless dictionaries in more detail.

3.1. *MorDebe database*

The most frequent use of dictionaries is to look up if a word exists—around 80% of all uses according to Oppentocht and Schutz (2003)—. Similarly, on language consultancy sites, such as *Ciberdúvidas da Língua Portuguesa*, questions on whether a specific word exists are more frequent than any other (around 25% of the questions in the case of Ciberdúvidas). Many common users consider the dictionary to be an exhaustive repertoire of all words of the language. Not only when playing word games such as Scrabble, but even in more day-to-day language use do they consider only words that are listed in the dictionary as correct words.

Despite the fact that too many people take dictionaries to be the official source of all words, they are never intended as complete word lists. In the words of Erin McKean, chief editor of the Oxford American dictionary: "the dictionary is not the word social register" (Google conference). Dictionaries include words because they are evidenced in reference corpora, because their inclusion in the dictionary is useful, and because there is space left for them in the dictionary. This means amongst other things that dictionaries do not only leave out word that are used very infrequently or not at all, but also leave out words whose form and meaning is fully predictable by the rules of grammar and morphology.

Because of the fact that dictionaries leave out predictable words, they are not always the ideal tool for looking up whether a word exists. Imagine a user who wants to know whether the word *duríssimo* (very hard) is a correct Portuguese word. The word is not in the dictionary, since it is the regular superlative of *duro* (hard), which means not merely a predictable form, but even a form that is taken to be inflectional, and hence even less likely to appear in the dictionary. Many dictionaries will provide a separate entry for the suffix *–issimo* itself, indicating that it is a formation element for the synthetic superlative, which can be applied to adjectives. But commonly no direct indication of the applicability of this suffix to the adjective *duro* can be found in the dictionary. The user is expected to deduce that *duríssimo* is a correct Portuguese word from the existence of the adjective and the regular suffix.

Furthermore, there is little if any information in the dictionary to distinguish the correct form *duríssimo* from the incorrect form **malíssimo*. That latter form would be the correct regular application of the superlative suffix to the adjective *mal* (bad), but the correct superlative for *mal* is the suppletive form *péssimo* (very bad). Many dictionaries will have a headword entry for *péssimo*, and many dictionaries even list *péssimo* in some form in the entry for *mal*. But there is no clear or easy indication whether or not the form **malíssimo* is correct or not—there are are other adjectives such as *pobre* (poor), for which both the irregular form (*paupérrimo*), and the regular form (*pobríssimo*) are correct, but *mal* is not one of them—. Therefore, the user would be much better server by a resource that would include all the correct forms *paupérrimo, pobríssimo,* and *duríssimo,* but leave out the incorrect form **malíssimo.*

The need for indications of the existence of predictable forms applies not only to the superlative forms, but to a whole range of forms that are commonly absent from dictionaries. In Portuguese these include at least the following: the regular adverbs on *–mente*, the diminutives on *–inho, -zinho,* and *–zito* for both nouns and adjectives, the aumentatives on *–zão*, the deverbal nouns on *–ção* and *–mento,* and the regular feminine forms of animate nouns. For the end user, it would be ideal if the dictionary would tell the distinction between the actual and the potential lexicon (Hoeksema 1985).

The core lexicon behind the Portal, called *MorDebe*, attempts to get closer to describing the actual lexicon by providing a full-form lexicon including the inflected forms of all the lexical

entries, as well as predictable derivations or compounds. It is a list of the common predictable and unpredictable citation forms and word-forms of the language, without their meaning definition. Since the purpose of MorDebe is not to explain words, it is irrelevant whether words are predictable or not: MorDebe lists words because they occur in reference resources, not because their occurrence is surprising or their meaning is in need of clarification.

Because MorDebe is a digital lexicon with no printed form, the greater number of entries a full-form lexicon requires is not problematic in the way it would be for a paper dictionary: there are virtually no space limitations. And since no meaning definitions have to be given, the additional time required for each additional entry is minimal as well. Because of the lack of need to generate description, there is also more time available to verify the orthography of all the entries more thoroughly, and adapting misconstrued entries will be easier. The fact that there are only headwords in the database will also make it easier to change the orthography of the lexicon when the recent spelling reform for Portuguese, which is currently under evaluation, will be implemented.

3.1.1. *Problems with a meaningless lexicon*

Despite the advantages for the user and the lexicographer described above, leaving out meaning definitions in the lexicon is not without problems. Firstly, many users have difficulties understanding the use of a database without meanings. Given that MorDebe has a feedback system—we get some impressions about how people perceive the web site, and the most common issue brought up by the users is exactly the lack of meanings—. The user perception will be described in more detail in chapter 4.

Secondly, and more significantly, without an indication of meaning, the lexicon will tell the user that the word he encounters is correctly spelled, but not if it is also the word he was trying to spell. A classic example is the word *aardbij* in Dutch, which is a correctly spelled word, yet much more likely to be a misspelling of the word *aardbei* (strawberry). Without the indication that an *aardbij* is a type of bee living underground, a lexicon containing that word will simply be misleading to the users—despite the fact that it provides only correct information—. This is the same problem as with spelling checkers: the larger the lexicon of a spelling checker, the larger the chances that a typographic error will be homographous with one of the words in the lexicon. And it is a problem that is shared by all the official vocabularies as well.

3.1.2. *Solutions in MorDebe*

The fact that many predictable forms are included in MorDebe means that there are more words in the database than there would be in a traditional dictionary, and hence more possibilities of undesirable homographs. But for the regular derivations included in Mordebe, there is a solution to this problem: although MorDebe itself does not provide meanings, the other OSLIN resources provide links for many of the predictable forms back to the word they are derived from or related to. This because many of the single-task dictionaries in OSLIN provide exactly the type of information that a derivation, or better *inherent inflection* form (Booij 2005), is derived from its root from. An example is the information in the deverbal noun dictionary described in the next section, which links all the regular (and irregular) deverbal nouns to their related verbs. OSLIN not only provides this information for the deverbal nouns, but for many other inherent inflection relations as well. As explained in chapter 2, the entry page for each word in MorDebe present all the related information in the various other resources in OSLIN, including this relation between inherent inflection forms and their base form. Therefore, most of the potential undesirable homographs in the set of morphologically predictable words are indicated as being inherent inflections in MorDebe, with a remission to their base form. Hence, they should not be confused with whichever word they might have otherwise been confused with.

This solution for "disambiguating" is in a sense making use of meaning definitions—not traditional, full-fledged meaning definitions, but semantic characterizations of the words nevertheless, and semantic characterizations that are external to the meaningless MorDebe database itself—. Also, it is a solution that only works for a limited selection of words—those that

happen to be inherently inflectional—. A more general solution to the problem of potential misinformation is to provide notes with all words that could be misperceived. For instance, in MorDebe there is a footnote with the word *microonda* (microwave) that it is a word from the domain of physics, to tell it apart from the plurale tantum *micro-ondas* (microwave-oven). If necessary, these footnotes could of course be full semantic definition, implying that semantic definitions are only given in those cases where their absence would lead to confusion. Although not an ideal solution, it at least solves the problem in a practical way in most cases.

Since MorDebe provides the complete inflectional paradigm for each lexical entry, it is also necessary to provide a distinguishing note with all homographs that have a different inflection, such as the English verb *to ring* which is irregular (rung) in its normal meaning, but regular (ringed) in its meaning of putting a ring around the leg of a bird. Cases like these are extremely rare in Portuguese though, the only known examples are the verbs *redar* (irregular when meaning "to give again", but regular in the sense "to catch with a net"), and *saber* (which has only one deviant form—the first person singular present indicative is *sei* for the verb "to know", and *saibo* for the verb "to have taste"—).

The solution of providing footnotes with problematic entries might not be the most elegant solution ever, but with it, the advantages of a meaningless lexicon outweigh the disadvantages: for the mere questions of correct spelling and word existence, a simple list of correctly spelled words, whether predictable or not, is more useful to the user than a general language dictionary. The advantages from the perspective of the lexicographer might seem insignificant: merely a decrease of the time required. But the selection of the basic lexicon is a task that often gets less attention in the process of making a dictionary than it deserves, whereas it might be one of the most criticized issues by the common user. Apart from academic criticism, it is uncommon to find negative comments about dictionaries based on the semantic definitions that they give. It is much more common to find objections against the inclusion of certain words, criticism against the spelling of adapted loanwords, and rejections of the inflected forms indicated in the dictionaries. A lexicographic work that concentrates solely on the selection of correct words with their appropriate spelling and inflection is hence advantageous for both the user and the lexicographer.

3.2. *Dictionary of deverbal nouns*

Dictionaries present a large number of deverbal event nouns (henceforth DEN)—nouns that express the abstract event related to the verb, such as the word *destruction* for the verb *destroy*—. The inclusion of these nouns in the dictionary is normally not because of their regular reading, which is predictable and hence by default not included in the dictionary. Their inclusion is due to the fact that many DEN have additional unpredictable readings aside from their regular DEN meaning, and the event reading is registered in a sense as a by-product.

The most typical entry in the dictionary for a DEN has an indication in the etymological field that the noun is derived from verb+suffix, and a meaning definition *action (or effect) of X,* where *X* is the verb in question. However, several alternatives to this pattern can be found:

1. The meaning definition uses a different phrasing instead of "*action (or effect) of*" such as *act of X*, *act or operation of X*, or *action or period of X*

2. If the deverbal noun existed in Latin, the noun is often etymologically marked as a direct loanword from Latin rather than a contemporary morphological derivation—as in the case of *operação* (operation), which comes from *operatio +onis* acording to DLPC.

3. The verb is defined in terms of the noun, even though the noun is (also) derived from the verb. An example of a reverse case: *exitar* is defined in GDLP (2004) as *provocar excitação em*, even though it is not a case of back formation.

4. The noun is defined not in terms of the verb, but rather in parallel with the verb—as in the example *igualação* (levelling), which is defined in GDLP (2004) not as *acto ou efeito de igualar* (action or effect of levelling), but rather as *acto ou efeito de tornal igual ou nivelar* (action or effect of making equal or level), where *igualar* (to level) has the meaning *tornar igual* (to make equal) and *nivelar* (to make level).

414

Although all these alternative definition, as well as the restriction selection criteria for deverbal nouns, are well motivated from the perspective of the general language dictionary, it makes the dictionary not necessarily the ideal tool for someone interested in finding the nominal form of any given verb. Firstly, since DEN are mostly included when they acquire an unpredictable meaning, many fully transparent nouns will be left out. That means that for transparent cases, the dictionary will contain no information about what the deverbal noun is. And although the noun might be transparent, there are at least two very productive deverbal suffixes in Portuguese: -*mento* and *–ção*. Many verbs accept either form, but for most verbs only one of the two is correct. Although looking at the noun it might be clear which verb it belongs to and what its meaning is, looking at the verb it is less predictable what the correct deverbal noun will be.

Secondly, even for the deverbal nouns that are included in the dictionary, it is the verb that is included in the definition of the noun, not the other way around. So any user interested in finding the noun for a given verb will have to guess what the deverbal noun might be before he can look it up. In most cases, the two forms will be close together in the dictionary, given the suffixing nature of regular deverbalization. But for several irregular cases, such as the noun *cessão* for the verb *ceder* (to cease), the noun is not that easy to find unless you know what it is.

Thirdly, even if the user manages to find the correct deverbal form, the reverse and parallel cases make it sometimes hard to judge whether the noun is indeed the denominal form of the verb. For all these reasons, the *Dicionário de Nomes Deverbais* on the *Portal* provides a complete lists of all the deverbal nouns for each verb that has one (or more). The advantages for the user interested in finding what the deverbal noun is for any given verb are obvious: the dedicated dictionary gives the desired information in a quick and simple way, whereas the general language dictionary is either lacking the information, or presenting it only in an indirect way. And since the presentation of the noun as the nominal form of a verb *is* in a way a semantic definition, the user experiences less problems with the absence of a semantic definition.

In the case of the dictionary of deverbal nouns, the advantages for the lexicographer are also more obvious: firstly, in the presentation of the dedicated dictionary, it is possible to include a preface explaining exactly what the purpose of the dictionary is, and what the entries in the dictionary intend to represent. On the Portal, the dictionary comes both with a small introduction, and with a more detailed description about the nature of the entries, and the selection criteria applied. Secondly, the choice of the definition becomes easier because only the standard definition is relevant: whether the noun can (also) express a period of time is irrelevant from the perspective of this dictionary; whether the noun is synchronically formed or adopted as a deverbal noun from Latin is also redundant information; and reverse or parallel definition should not be necessary. The choice for such a definition is mostly driven by the unfamiliarity of the verb, and since the dictionary of deverbal nouns does not intend to give a full semantic description, it is less relevant how familiar the verb is. Thirdly, the smaller scope of the dictionary makes it possible to complete it within a relatively short amount of time, with the possibility of using computer generated tools for finding all candidate DEN forms, making it much easier to apply the same set of criteria throughout the dictionary.

However, leaving out the definitions also has drawbacks in the case of the deverbal nouns. Firstly, without definitions it is not possible to indicate the potential usage differences between the different deverbal nouns of the same verb. In Portuguese, there can be up to seven correct deverbal nouns for the same verb: for instance *arrancadura, arrancada, arrancamento, arranca, arrancadela, arranco,* and *arranque* are all correct DEN for the verb *arrancar* (to pull off). Although the differences are small, these forms are not fully interchangeable. For instance *arrancadela* typically indicates a light way of pulling off. In other cases, there can be usage restriction on some of the deverbal forms, which can be more or less formal, regionally delimited, or simply infrequently used, as in the case of *abafeira* for the verb *abafar* (to choke).

Secondly, there are cases, although rare, in which the deverbal noun is different depending on the meaning of the verb. For instance, the noun *embicadura* is only the correct deverbal noun of

embicar in the sense of "heading into port". In all other senses of the verb, only *embicação* can be used. Without semantic definition for either the noun or the verb, there is no way of indicating these semantical restrictions.

As for the first problem, the main way in which general language dictionaries can deal with such differences is by means of usage labels. Usage labels are not part of the semantic definition, but in a sense external markers associated with the word. Usage labels could be applied to the deverbal noun dictionary in the Portal in the same way as they are used in general language dictionaries: by providing the entries in the deverbal noun dictionary with usage labels, the same information can be given (although the Portal does not currently provide such usage labels). Besides, in the case of the deverbal noun dictionary, the interpretation of the usage labels is more transparent than they are in general language dictioanries: *uncommon* will always be interpreted with respect to the alternative deverbal nouns for the same verb—it is not an absolute statement, but a statement about the relative status of a specific deverbal form with respect to the alternative forms for the same verb.

With respect to the second problem, it is very uncommon for deverbal nouns to be restricted to specific senses of the verb. For the rare occasions in which it does happen, it is best resolved with a footnote. In the Portal, however, a more radical solution was adopted for those cases. Since the deverbal noun *embicadura* is not really related to the verb *embicar*, but to a word-sense of the verb *embicar*, the deverbal noun was not included in the dictionary. Whichever of these solutions is adopted, an explanatory note for these marginal cases is always recommendable.

3.3. *Dictionary of gentiles and toponyms*

Adjectives or nouns indicating things or people from a specific place, such as for instance *Albanian* for the country of Albania, are often referred to as *gentiles*. There are at least two problems with the treatment of gentiles in dictionaries. Firstly, there is an incredible amount of gentiles that are in actual use—there are gentiles not only for continents (*Australian*) and countries (*Belgian*), but also for smaller localities like villages and neighbourhoods—. The sheer amount of them, and their relative predictability, makes it necessary to have strict selection criteria for the inclusion of gentiles. The fact that many of these gentiles are used with a relatively high frequency near the place they relate to makes it necessary to have a much higher cut-off point for frequency and to apply a different interpretation for the term *regionalism* for these words.

Secondly, many dictionaries are (largely) etymology driven when it comes to separating homonyms from polysemous words. When two cities have the same name, as is not uncommon, their gentiles also tend to be identical. But since the gentiles are derived from distinct yet homographous toponyms, the etymology of the gentiles is by definition distinct as well. In order to save the etymological treatement of word identity, it would therefore be necessary to identify all these gentiles as distinct lexical entries, as is indeed sometimes done: the Houiass dictionary for Brazilian Portuguese lists a staggering total of 22 entries for the word *santanense* (from Santana), and 21 entries for the word *barrense* (from Barros).

Most gentiles are quite regular and predictable: in Portuguese, the most common way of forming gentiles is by adding the suffix *–ense* to the toponym, as in *Madalena – madalenense*. But there are other toponyms for which the gentile is less predictable. The parts of a multi-word toponym can be inversed or (partially) translated as in *Aguas Calientes – hidrocálido* (hot water). The gentile can relate to an ancient name of the place *Santarem – escalabitano*. Or the name can just be colloquial—as in *Rio de Janeiro— carioca* (for the city) or *fluviense* (for the state). And there are toponyms for which there are various alternative gentile, such as for *Lisboa* (Lisbon, Portugal)—which can be either of the following: *lisbonense, lisboeta, olisiponense,* and the less common *lisboano, lisboês, lisbonês,* and *lisbonino*.

Dictionaries commonly define gentiles in terms of the toponym they belong to—as for instance in the LDOCE entry for *Dutch*: "of the people, country, or language of the Netherlands (Holland)"—. But given the existence of irregular gentiles, it is not always easy for the user of a general language dictionary to find the appropriate gentile for a given toponym. That is why the *Dicionário de*

Gentílicos e Topónimos of the Portal gives a full description of a large selection of toponyms and their gentiles, either by alphabetic order by toponym or by gentile, or by means of a search.
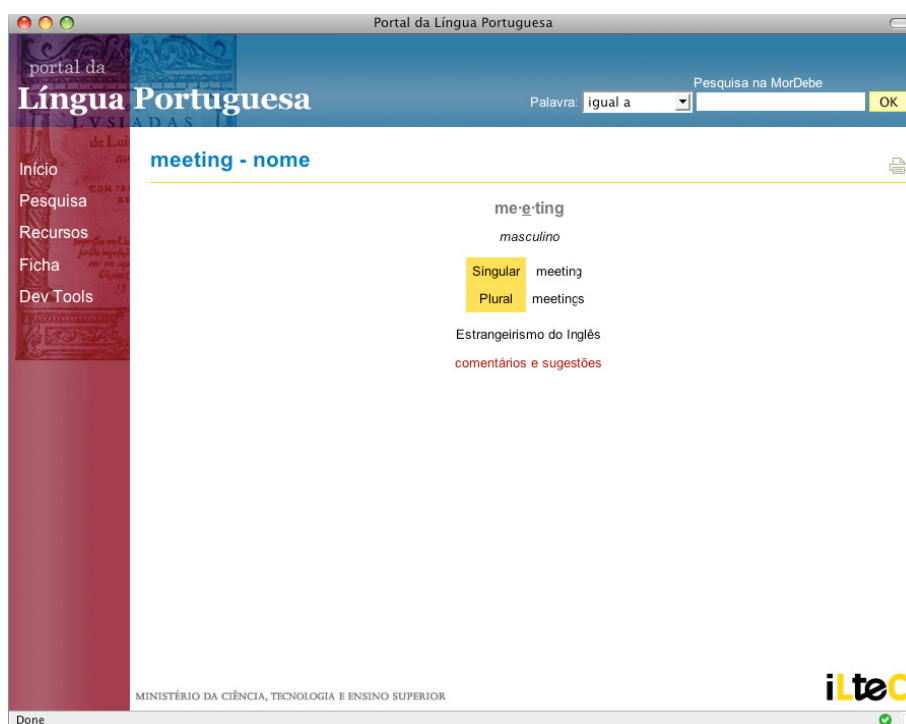

Figure 2. The alphabetic representation of the gentiles dictionary of the Portal

Gentiles are in a sense the most obvious candidate for a meaningless dictionary, which is also why they are quite commonly included as separate lists in general language dictionaries, such as, for instance, DLPC and LDOCE, as mentioned before. Since toponyms are not commonly included in general language dictionaries, the choice for a separate list, or better for a single-task meaningless dictionary, is very obvious. In a separate dictionary for gentiles, the amount of gentiles becomes less problematic, and although it remains important to apply strict criteria in a single-task dictionary, these criteria do not need to be as restrictive as they have to be in a general language dictionary.

Even in the case of gentiles, the creation of a single-task dictionary leads to problems, although less prominently because of the lack of semantic definitions. The most semantically related problem is the issue whether derogative nouns, such as *yankee* for things or people for the United States, should be included in the dictionary. Whether or not to do so is a design choice, which should be carefully explained in the preface of the dictionary. The gentiles dictionary of the Portal opted not to include such nouns, although it is not always easy to distinguish common gentiles from derogatory ones—we have received several responses indicated a derogatory value for certain toponyms from Cape Verde, despite the fact that none of the reference resources acknowledge such a usage label.

More problematic for a full-scale dictionary of gentiles is the lack of consensus surrounded the gentiles for smaller places. The gentiles indicated in existing dictionaries often go directly against the terms used by the people living in the village in question, different generations or different neighbourhoods from the same village often seem to use a different adjective for their own village, and in many cases, political issues play an important role in the acceptability of a certain designation. But all these problems are not due to the absence of meaning in the gentiles dictionary on the Portal, and are not indications against the usefulness of a meaningless dictionary for this purpose, but rather a clear indication for the dire need for a reference work of this type, despite its inevitable controversiality.

4. User responses

As stated in chapter 2, the Portal features the possibility for users to provide feedback on potential errors they encounter, in order to help improve the quality of the databases. Although sporadically, this feedback form has been useful for the improvement of some imperfectly treated entries. But the fact that there is a feedback form, means that users provide various other types of comments as well. Apart from the spam messages which inevitably show up on any Internet form, and the very odd questions such as a question by a person what to do now that she had had unprotected intercourse, this feedback gives some indications as to how the end users appreciate the content of the web site.

Despite the fact that the page on which users can enter their comments very clearly states that the Portal does not provide semantic information, that we will not answer questions regarding semantics, plus links to online dictionaries and the *Ciberdúvidas* site where people *can* get information about meaning, the most common feedback by far consists of question about the meaning of a specific word. Also, whenever people provide suggestions as to what might be a useful addition to the site, the suggestion is without failure that they would like to see semantic definitions. And whenever we receive hate mail, although that fortunately does not happen very often, people are almost invariably yelling that the site does not make sense because it does not even say what the word they were looking at means.

However, the Portal has a very steady base of users, and most users that start using the resources on the site keep coming back. This means that although the principle of specialized dictionaries without meaning might be upsetting incidental visitors, who are expecting a free, full-scale, all-encompassing general language dictionary, the meaningless dictionaries do in fact serve the purpose they were intended for: to provide users with a specific question with exactly the type of information they are looking for. The general MorDebe lexicon seems to be more problematic than the more specific dictionaries: there are hardly any negative comments about the dictionary of deverbal nouns or the dictionary of gentiles, despite their quite frequent use. Visitors can get very involved in their demand to add, modify, or delete some information concerning the place they happen to be living in, but overall, the reception of both of the more specific dictionaries appears to be very positive.

5. Conclusion

As shown by the examples in this paper, small single-purpose dictionaries without meaning definitions are better suiting the needs of the dictionary than a general language dictionary in a wide range of cases. Apart from the advantages for the user, building a single-purpose dictionary without meaning definitions can be much more feasibly done with small financial and human resources, while still maintaining a high level of quality in terms of completeness, correctness, and coherence.

However, both for the dictionary maker and for the end user, the deployment of meaningless dictionaries carries several problems as well: the lexicographer cannot rely on the meaning definitions to make small and subtle distinctions, the end user is still expecting to see the word meaning posted alongside with the information he is looking for, and in some cases the lack of meanings can lead to confusing situations. But especially when combined with a careful use of explanatory notes in problematic cases, single-task meaningless dictionaries can provide several types of information traditionally covered in the dictionary in a better, cheaper, and more coherent way than the general language dictionaries do.

References

Booij, G. (2005). "Inflection and derivation". In Brown, K. (ed.). *Encyclopedia of Language and Linguistics*. Oxford: Elsevier.

[DLPC]. *Dicionário da Língua Porguesa Comtemporânea*. Lisbon: Verbo, 2001

[GDLP]. *Grande Dicionário da Língua Portuguesa*. Porto: Porto Editora, 2004

Hanks, P. (2000). "Do Word Meanings Exist?". *Computers and the Humanities* 34 (1-2). 205-215.

Hoeksma, J. (1985). *Categorial Morphology*. New York: Garland.

Houaiss, A. (2001). *Dicionário Houaiss da Língua Portuguesa*. Rio de Janeiro: Objetiva.

[HWNT]. *Herziene Woordenlijst van de Nederlandse Taal*. Den Haag: SDU Uitgeverij, 1990

Janssen, M. (2005a). "Open Source Lexical Information Network". In *Third International Workshop on Generative Approaches to the Lexicon*. Geneva, Switzerland.

Janssen, M. (2005b). "Between Derivation and Inflection". In *East West Encounter: second international conference on Meaning - Text Theory*. Moscow, Russia.

Kilgarriff, A. (1997). "I Don't Believe in Word Senses". *Computers and the Humanities* 31(2). 91-113.

[LDOCE]. *Longman Dictionary of Contemporary English*. Essex: Longman, 1987.

Oppentocht, L. and Schutz, R. (2003). "Developments in Electronic Dictionary Design". In Sterkenburg, P. van (ed.) *A Practical Guide to Lexicography*. Amsterdam: John Benjamins Publishing Company. 215-227.